### Title

LipidBlast templates as flexible tools for creating new in-silico tandem mass spectral libraries.




### Journal

Analytical chemistry, 86(22)

### ISSN

0003-2700



### Authors

Kind, Tobias

Okazaki, Yozo

Saito, Kazuki

et al.



### Publication Date

2014-11-01

### DOI

10.1021/ac502511a

Peer reviewed

analytical chemistry

Technical Note

pubs.acs.org/ac

# LipidBlast Templates As Flexible Tools for Creating New in-Silico Tandem Mass Spectral Libraries

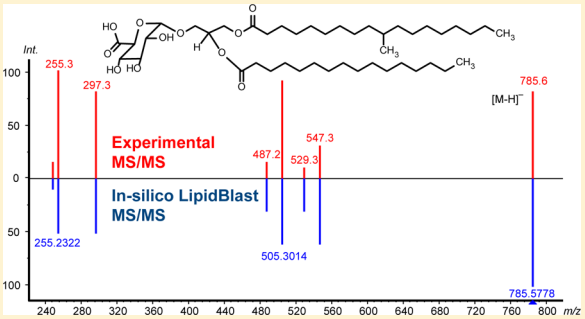
Tobias Kind,*[,†] Yozo Okazaki,[‡] Kazuki Saito,[‡,§] and Oliver Fiehn*[,†]

[†]West Coast Metabolomics Center, University of California Davis, Davis, California 95616, United States

[‡]RIKEN Center for Sustainable Resource Science, Metabolomics Research Group, Yokohama 230-0045, Japan

[§]Graduate School of Pharmaceutical Sciences, Chiba University, Chiba 260-8675, Japan

**ABSTRACT:** Tandem mass spectral libraries (MS/MS) are usually built by acquiring experimentally measured mass spectra from chemical reference compounds. We here show the versatility of in-silico or computer generated tandem mass spectra that are directly obtained from compound structures. We use the freely available LipidBlast development software to generate 15 000 MS/MS spectra of the glucuronosyldiacylglycerol (GlcADG) lipid class, recently discovered for the first time in plants. The generation of such an in-silico MS/MS library for positive and negative ionization mode took 5 h development time, including the validation of the obtained mass spectra. Such libraries allow for high-throughput annotations of previously unknown glycolipids. The publicly available LipidBlast templates are universally applicable for the development of MS/MS libraries for novel lipid classes.




De novo structure elucidation of novel compounds require multiple separation and purification steps as well as the inclusion of complex analytical methods such as liquid chromatography coupled to high-resolution tandem mass spectrometry (LC−MS/MS) and nuclear magnetic resonance spectroscopy (NMR). Such an approach has been shown recently for the discovery of glucuronosyldiacylglycerol lipids (GlcADG) from phosphorus depleted plants.[1] An alternative strategy for compound annotation is matching experimental MS/MS spectra to in-silico predicted MS/MS spectra as it has been shown for complex lipids using LipidBlast,[2] demonstrating the applicability of such in-silico libraries for 50 different types of low resolution and high resolution mass spectrometers. While in-silico libraries like LipidBlast use heuristic modeling of tandem mass spectra,[2] prediction of mass fragmentation patterns using first-principle methods has been described recently by applying quantum chemical calculations for the generation of $MS^1$ electron ionization (EI) spectra.[3] Such principles might be transferable to collision-induced tandem mass spectra in the future. LipidBlast encompasses over 200 000 tandem mass spectra and covers 25 lipid classes and also provides development tools using Microsoft Excel template files[4] that can be utilized to create in-silico MS/MS libraries for novel lipid classes and multiple adduct ions. We have previously shown the versatility of such an approach for the identification of mammalian stem cell lipids,[5] *Chlamydomonas reinhardtii* and *Chlorella minutissima* algal lipid research[6,7] and the identification of lipids in the flagellate protists *Euglena gracilis*.[8] Using development templates in standard software has several advantages. It enables (1) addition of novel lipid classes that had not been covered in the original LipidBlast survey, (2) addition of different acyl-chain lengths or degrees of unsaturation that had not been covered by LipidBlast, and (3) fast distribution of such open tandem mass spectral libraries across laboratories. We here exemplify how a new in-silico library can be developed including 15 000 MS/MS spectra for positive and negative ionization mode in less than 5 h development and validation time after experimental discovery of a single new plant lipid.

## METHODS

The development of new LipidBlast MS/MS spectra requires a development sheet (Microsoft Excel) and a Visual Basic for Applications software (Microsoft Excel VBA) that contains programmatic code to export the MS/MS spectra. Extensive method details can be found in the original LipidBlast publication.[2] We concisely describe the general approach here. The creation of new in-silico MS/MS libraries requires sample MS/MS spectra for development and a series of MS/MS spectra for validation. Spectra were taken from original experimental measurements and the publication itself.[1] These tandem mass spectra were used for developing the library. The spectra were acquired on a hybrid ion trap/time-of-flight mass spectrometer with electrospray ionization (ESI-IT-TOF-MS). Additional supplemental raw MS/MS spectra from the same instrument were used for validation. All spectra are freely available for download from our Web site. Instead of a completely independent development, we simply copied an

**Received:** July 8, 2014
**Accepted:** October 23, 2014
**Published:** October 23, 2014

   dx.doi.org/10.1021/ac502511a | *Anal. Chem.* 2014, 86, 11024−11027

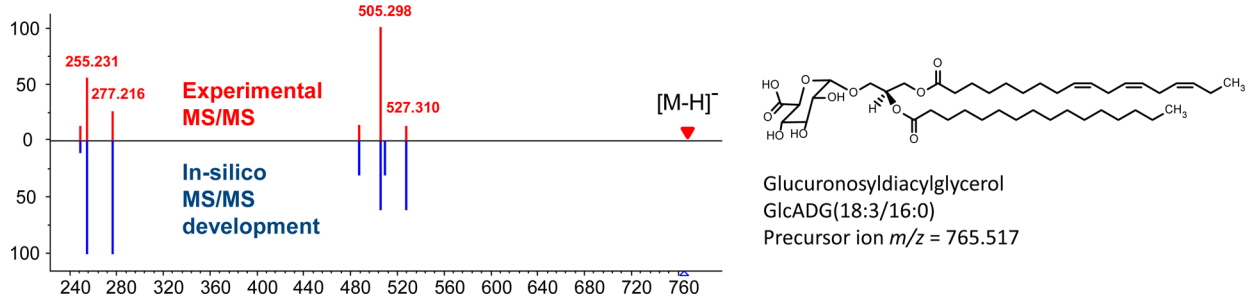



**Figure 1.** Single published tandem mass spectrum from a novel plant lipid class glucuronosyldiacylglycerol (GlcADG) was used for development of 5 000 related lipids. Using the LipidBlast templates in-silico MS/MS spectra with different acyl chain lengths and degrees of unsaturation were modeled. The tandem spectrum of GlcADG(18:3/16:0) shown here is observed in negative ionization mode precursor *m/z* 765.51529 Da.

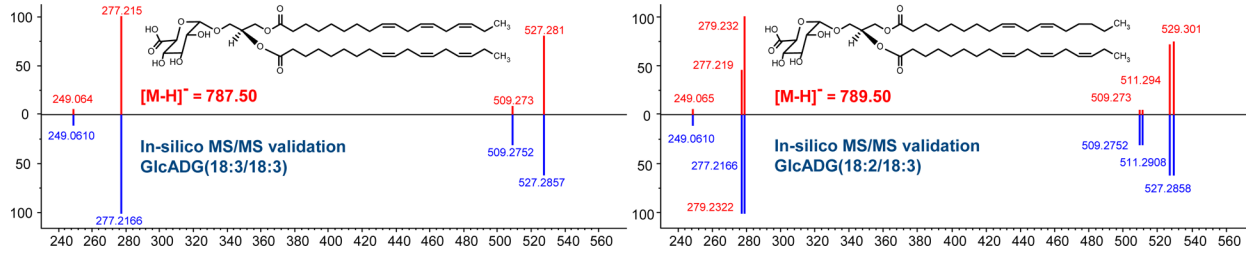



**Figure 2.** Additional published tandem mass spectra were used for validation of the novel LipidBlast library. All in-silico MS/MS spectra were created using the freely available LipidBlast templates. In top panels (red) the experimental MS/MS spectra are given, in lower panels (depicted in blue), predicted MS/MS spectra validated this approach. These spectra can be used for high-throughput annotations of lipids.

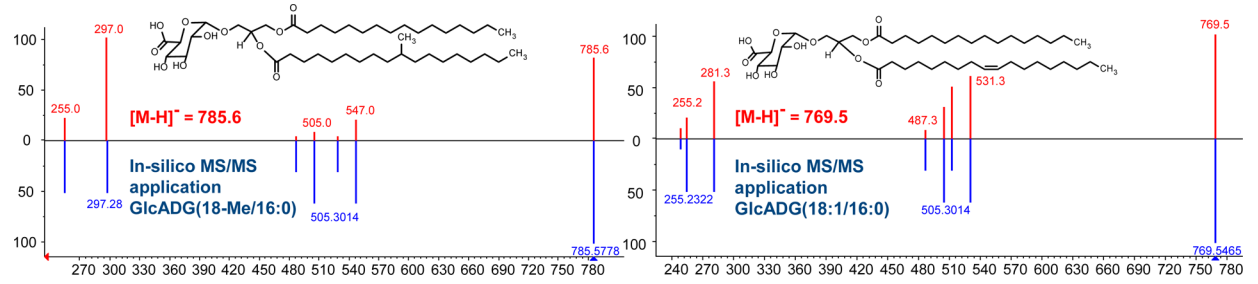



**Figure 3.** In-silico library can be used for assignment of MS/MS spectra from different platforms. Left panel, MS/MS from Finnigan-MAT TSQ70 with FAB ionization of a *Mycobacterium smegmatis* glycolipid[14] and right panel, MS/MS from Thermo-Finnigan LTQ-FT-MS.[11] The in-silico spectra shown here are low-CID voltage spectra with abundant precursor ions. Experimental spectra are depicted on top (red) and in-silico MS/MS spectra are shown on the bottom (blue).

existing member from the LipidBlast development sheet (MSMS-prediction-distribute-v49.xls)[4] and chose a lipid class (sulfoquinovosyldiacylglycerol, SQDG) similar to the novel GlcADG lipid class as a starter template. In-silico structures were generated by replacing the SQDG core structure with GlcADG in the SMILES (Simplified Molecular Input Line Entry System) structure codes using text based find-and-replace in Microsoft Excel. The resulting SMILES codes were used to calculate a series of required molecular properties using the ChemAxon cxcalc command line tools.[9] The properties included were accurate mass, octanol−water partition coefficient (log *P*), and the InChIKey. The values in the copied SQDG Microsoft Excel template sheet were subsequently adjusted for the novel GlcADG adduct ions as well as observed GlcADG fragmentations. The associated VBA code was modified to allow the export of the GlcADG lipids. Tandem mass spectrometry MSP files were converted by the LIB2NIST program, and the resulting libraries were copied as a subdirectory into the NIST MS Search program. The libraries were then used to further validate MS/MS spectra using the NIST MS Search program and for batch-wise comparison using the NIST MSPepSearch program.[2,4]

## RESULTS AND DISCUSSION

We created a total of 15 000 novel in-silico MS/MS spectra for the glucuronosyldiacylglycerol lipid class using the LipidBlast development templates. A total of 5000 tandem mass spectra were modeled for positive $[M + NH_4]^+$ ionization mode and 10 000 MS/MS spectra for negative ionization mode $[M - H]^-$. The negative ionization mode numbers are twice the size because they cover spectra for low-CID (collision-induced dissociation) and high-CID voltage mode. Lipids with acyl carbon chain lengths $C_2$ to $C_{26}$ and degrees of unsaturation with double bond counts of 0−6 are included. The peak fragments and their individual abundances were modeled according to a reference spectrum obtained from an ion trap/time-of-flight mass spectrometer (see Figure 1). This approach is feasible because lipids follow very consistent fragmentation rules. For MS/MS library search we used accurate mass precursor search with subsequent product ion matching. The details of the matching procedure are outlined in the original


dx.doi.org/10.1021/ac502511a | *Anal. Chem.* 2014, 86, 11024−11027

LipidBlast paper.[2] Short, the precursor filter removes many false candidates that fall outside a given mass window. The subsequent product ion matching algorithm uses traditional similarity scoring of remaining candidates. Reverse search scores can be used in case of impurities or nonexplained peaks.[10]

We validated the negative ionization mode in-silico MS/MS spectra with four experimentally obtained tandem mass spectra from the same class (see Figure 2). These experimental spectra matched the in-silico generated spectra, with one spectrum (GlcADG 36:4) generating multiple assignments, due to nonresolved ion peaks and overlapping product ions. In the case of overlapping or not completely resolved peaks by liquid chromatography as shown for GlcADG 36:4, lower hit scores with more ambiguous compound annotations are obtained. In order to increase hit scores we additionally modeled spectra for low-CID and high-CID voltage mode. In low-CID voltage the precursor ion has a higher peak intensity because it is not completely fragmented. In high-CID voltage mode the precursor ion disappears due to complete fragmentation and the fatty acyl intensities highly increase. The CID voltage specific modeling allows for analysis of experimental spectra from a wider range of instruments such as triple quadrupole or Fourier transform (FT) mass spectrometers (see Figure 3). A final validation and application step was performed on tandem mass spectra obtained from authentic reference standards synthesized by Cao and Williams.[11] The paper also discussed specific product ion ratios for $[M - H{-}sn2 + H_2O]^-$ and for $[M - H{-}sn1 + H_2O]^-$ that can lead to the correct positional assignment of sn1 and sn2 fatty acyls. Such total synthesis approaches and detailed CID investigations with different ionization voltages will be extremely valuable for assignments of regioisomers in future versions of LipidBlast. Currently LipidBlast libraries cannot annotate stereochemistry, regiospecificity, and position of double bonds correctly. Also a number of bacterial fatty acids such as cyclic-, prenyl-, and epoxy fatty acids are not yet included. However, the lipid class as well as the total carbon and degree of unsaturation of each of the fatty acyl chains can be correctly annotated. The positive ionization mode $[M + NH_4]^+$ spectra were developed in a similar way and validated with two independent MS/MS spectra. The positive ion mode spectra are specific for hybrid ion trap/time-of-flight mass spectrometers. No voltage optimization has been performed due to the lack of additional reference spectra.

Our publicly available GlcADG glycolipid library has direct translational aspects that go beyond plant lipid research.[12,13] Glucuronidated glycerolipids occur across several domains of life or phylogenetic branches. Glucoronidyl (glucuronosyl) lipids containing tuberculostearic acid (C18-methyl) were found and analyzed in mycobacterial lipid extracts[11,14,15] from *Mycobacterium smegmatis* (see left panel of Figure 3) and other species. Recent research discussed glycolipid antigen activity and the preferred binding of natural killer cells (Va10 NKT) toward glucuronosyl diacylglycerol lipids.[16] It has to be mentioned that in the current GlcADG LipidBlast library the C18-methyl group (tuberculostearic acid) is not directly assigned but rather annotated as C19 fatty acid. Glycolipids with similar structures were also observed in Gram-negative bacteria, *Pseudomonas diminuta*,[17,18] *Hyphomonas jannaschiana*,[19] *Agrobacterium tumefaciens*,[20] and Gram-positive bacteria such as *Corynebacterium glutamicum*.[21] GlcADG related glycolipids were also found in the fungus *Aspergillus fumigatus*.[22] Diacylglyceryl-alpha-D-glucuronide algal lipids have been found in *Pavlova lutheri* algae.[23,24] The GlcADG lyso-forms (one acyl chain) as well as ether analogues (plasmenyl, plasmanyl) have been described in the literature for use as lipid haptens[25] and synthesized for membrane property estimations,[26] but no evidence has been found that they exist in nature yet.

Most of the publications did not report MS/MS spectra in the past. Subsequently such spectra could not be accumulated in large electronic mass spectral databases such as Wiley MSforID,[27] ReSpect,[28] MassBank,[29] NIST,[30] or Metlin.[31] We close that gap with our publicly available in-silico MS/MS library, enabling future research groups to perform high-throughput analysis of complex glycolipid mixtures by simply extending and using LipidBlast.

The fast development of in-silico MS/MS spectra using the LipidBlast Excel templates shows the versatility and broad application domains of our LipidBlast software. The developed libraries and new templates are freely provided for commercial and noncommercial reuse with a Creative Commons-By Attribution (CC-BY) license and can be found under http://fiehnlab.ucdavis.edu/projects/LipidBlast.

## AUTHOR INFORMATION


### Corresponding Authors

*E-mail: tkind@ucdavis.edu.

*E-mail: ofiehn@ucdavis.edu.


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank ChemAxon for a free research license of the Marvin cheminformatics tools. We thank the NIST Mass Spectrometry (US) group for providing the freely available NIST MS Search GUI and MSPepSearch program. Data from Cao et al.[11] was used to generate the experimental spectrum for the table of contents/abstract graphic. Funding for T.K. and O.F. was supported by NSF Grant MCB 1139644 and NIH Grant U24 DK097154. Funding for K.S. was supported by JST SICORP (Metabolomics for a Low Carbon Society).

## REFERENCES


(1) Okazaki, Y.; Otsuki, H.; Narisawa, T.; Kobayashi, M.; Sawai, S.; Kamide, Y.; Kusano, M.; Aoki, T.; Hirai, M. Y.; Saito, K. *Nat. Commun.* **2013**, *4*, 1510.

(2) Kind, T.; Liu, K.-H.; Lee, D. Y.; DeFelice, B.; Meissen, J. K.; Fiehn, O. *Nat. Meth* **2013**, *10*, 755−758.

(3) Grimme, S. *Angew. Chem., Int. Ed.* **2013**, *52*, 6306−6312.

(4) Kind, T. LipidBlast Development Software. http://fiehnlab.ucdavis.edu/projects/LipidBlast, 2013.

(5) Meissen, J. K.; Yuen, B. T.; Kind, T.; Riggs, J. W.; Barupal, D. K.; Knoepfler, P. S.; Fiehn, O. *PLoS One* **2012**, *7*, e46770.

(6) Kind, T.; Meissen, J. K.; Yang, D.; Nocito, F.; Vaniya, A.; Cheng, Y.-S.; VanderGheynst, J. S.; Fiehn, O. *J. Chromatogr., A* **2012**, *1244*, 139−147.

(7) Dawei, Y.; Xiaolei, F.; Kind, T.; Fiehn, O.; Rongbo, G. *Acta Chim. Sin.* **2013**, *71*, 663−669.

(8) Ogawa, T.; Furuhashi, T.; Okazawa, A.; Nakai, R.; Nakazawa, M.; Kind, T.; Fiehn, O.; Kanaya, S.; Arita, M.; Ohta, D. *Biosci., Biotechnol., Biochem.* **2014**, *78*, 14−18.

(9) ChemAxon. Marvin Calculator Plugins; Marvin v5.9.9; http://www.chemaxon.com, 2012.

(10) Stein, S. E. *Anal. Chem.* **2012**, *84*, 7274−7282.

(11) Cao, B.; Chen, X.; Yamaryo-Botte, Y.; Richardson, M. B.; Martin, K. L.; Khairallah, G. N.; Rupasinghe, T. W. T.; O'Flaherty, R.

M.; O'Hair, R. A. J.; Ralton, J. E.; Crellin, P. K.; Coppel, R. L.; McConville, M. J.; Williams, S. J. *J. Org. Chem.* **2013**, *78*, 2175−2190.
(12) Hölzl, G.; Dörmann, P. *Prog. Lipid Res.* **2007**, *46*, 225−243.
(13) Ekins, S.; Williams, A. J. *Tuberculosis* **2014**, *94*, 183−185.
(14) Wolucka, B. A.; McNeil, M. R.; Kalbe, L.; Cocito, C.; Brennan, P. J. *Biochim. Biophys. Acta, Lipids Lipid Metab.* **1993**, *1170*, 131−136.
(15) Layre, E.; Sweet, L.; Hong, S.; Madigan, C. A.; Desjardins, D.; Young, D. C.; Cheng, T.-Y.; Annand, J. W.; Kim, K.; Shamputa, I. C. *Chem. Biol.* **2011**, *18*, 1537−1549.
(16) Uldrich, A. P.; Patel, O.; Cameron, G.; Pellicci, D. G.; Day, E. B.; Sullivan, L. C.; Kyparissoudis, K.; Kjer-Nielsen, L.; Vivian, J. P.; Cao, B. *Nat. Immunol.* **2011**, *12*, 616−623.
(17) Batrakov, S. G.; Il'ina, E. F.; Rozynov, B. V.; Bergel'son, L. D. *Russ. Chem. Bull.* **1975**, *24*, 85−91.
(18) Wilkinson, S. *Biochim. Biophys. Acta, Lipids Lipid Metab.* **1969**, *187*, 492−500.
(19) Batrakov, S. G.; Nikitin, D. I.; Pitryuk, I. A. *Biochim. Biophys. Acta, Lipids Lipid Metab.* **1996**, *1303*, 39−46.
(20) Semeniuk, A.; Sohlenkamp, C.; Duda, K.; Hölzl, G. *J. Biol. Chem.* **2014**, *289*, 10104−10114.
(21) Tatituri, R. V. V; Illarionov, P. A.; Dover, L. G.; Nigou, J.; Gilleron, M.; Hitchen, P.; Krumbach, K.; Morris, H. R.; Spencer, N.; Dell, A.; Eggeling, L.; Besra, G. S. *J. Biol. Chem.* **2007**, *282*, 4561−4572.
(22) Fontaine, T.; Lamarre, C.; Simenel, C.; Lambou, K.; Coddeville, B.; Delepierre, M.; Latgé, J.-P. *Carbohydr. Res.* **2009**, *344*, 1960−1967.
(23) Vishwanath, B.; Eichenberger, W.; Frey, F.; Frey, B. *Biochem. J.* **1996**, *320*, 93−99.
(24) Eichenberger, W.; Gribi, C. *Phytochemistry* **1997**, *45*, 1561−1567.
(25) Coulon-Morelec, M. J. *Bull. Soc. Chim. Biol.* **1967**, *49*, 825−840.
(26) Koynova, R.; Tenchov, B.; Kuttenreich, H.; Hinz, H. *Biochemistry* **1993**, *32*, 12437−12445.
(27) Oberacher, H.; Whitley, G.; Berger, B. *J. Mass Spectrom.* **2013**, *48*, 487−496.
(28) Sawada, Y.; Nakabayashi, R.; Yamada, Y.; Suzuki, M.; Sato, M.; Sakata, A.; Akiyama, K.; Sakurai, T.; Matsuda, F.; Aoki, T. *Phytochemistry* **2012**, *82*, 38−45.
(29) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K. *J. Mass Spectrom.* **2010**, *45*, 703−714.
(30) Yang, X.; Neta, P.; Stein, S. E. *Anal. Chem.* **2014**, *86*, 6393−6400.
(31) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. *Ther. Drug Monit.* **2005**, *27* (6), 747−751.